

A differential evolution optimized hybrid XGBoost for accurate carbon emission prediction

Lingfei Zhang^a , Gang Lu^b, Xiaoqing Yan^b, Peng Xia^b, Zhong Chen^c, Di Wu^{d,*}

^a Hanhong College, Southwest University, Chongqing, 400715, China

^b Energy Strategy and Planning Research Department, State Grid Energy Research Institute Co., Ltd, Beijing, 102209, China

^c School of Computing, Southern Illinois University, Carbondale, IL, 62901, USA

^d College of Computer and Information Science, Southwest University, Chongqing, 400715, China

ARTICLE INFO

Keywords:

Carbon emissions prediction
Hybrid deep learning
Model integration
Feature engineering
Correlation analysis

ABSTRACT

Predicting carbon emissions is essential for combating climate change and supporting green development. Carbon emissions are influenced by complex factors, such as economy, population and new energy generation. Traditional methods struggle with these uncertainties, while machine learning offers data-driven solutions. However, some models lack data selection strategies, resulting in the neglect of critical features. To tackle this issue, this paper proposes a Differential Evolution Optimized Hybrid XGBoost (DEOH-XGBoost) approach. DEOH-XGBoost includes three main components: feature engineering, model construction, and model integration. First, in each correlation analysis, features are selected through fuzzy membership functions. Second, XGBoost-based models are constructed on each feature set to predict separately. Third, the models are integrated by a differential evolution optimized weighting strategy. As such, DEOH-XGBoost effectively uncovers the intrinsic connections between multi-type data to achieve accurate carbon emission prediction. Extensive experiments demonstrate that our DEOH-XGBoost has significantly better prediction accuracy than related state-of-the-art methods. Our source code and datasets can be found at the following link: <https://github.com/lingfei0804/DEOHXGBOOST>.

1. Introduction

In the context of the development of industry and the over-exploitation of fossil fuels, the issue of excessive carbon emissions has become a major concern for human health (Ren et al., 2024). The industrial application of fossil fuels contributes greatly to these emissions. This emphasizes the urgency of achieving net-zero carbon emissions (Chen et al., 2024). Reducing carbon emissions has therefore become a common goal of global development (Aboagye and Sharifi, 2023). The control of industrial carbon emissions highlights the importance of public participation as an informal regulatory mechanism, on par with formal government environmental regulation (Wu et al., 2025). Accurate monitoring of industrial carbon emissions is essential for effective policy making and raising citizen awareness. In addition, relevant departments use carbon emission forecasts to ensure precise monitoring and control in the industrial sector. Therefore, accurately forecasting carbon emissions is a very important task.

First, vector regression, ARIMA, exponential smoothing, grey

prediction, and principal component analysis have become effective tools for energy forecasting. The second-order exponential smoothing model of vector regression combined data has achieved ideal results in predicting Chinas coal, oil, natural gas, and primary electricity demand (Rao et al., 2023). In addition, the ARIMA modeling, optimal forecasting technique modeling, and a decision support system based on fuzzy mathematics were applied to predict the catch of anchovies caught in a given port (Thessaloniki, Greece) during the period 1979–2000 and the total catch of cod and bonito during the period 1982–2000 (Koutroumanidis et al., 2006). In addition, the electricity demand forecasting model constructed using triple exponential smoothing (Holt-Winters) and seasonal autoregressive integrated moving average (SARIMA) produced small MAPE values (Shiwakoti et al., 2023). Subsequently, the model constructed using the nonlinear discrete grey power method performed well in predicting Chinas carbon dioxide emissions (Ding et al., 2020). Despite this, statistical learning methods usually rely on prior assumptions and feature engineering. This makes it challenging to effectively capture complex nonlinear relationships.

* Corresponding author.

E-mail address: wudi1986@swu.edu.cn (D. Wu).

Afterwards, the use of predictive models like transformers, XGBoost, Bayesian Networks, and Gaussian Processes surged in machine learning. For instance, a drought prediction transformer model was developed for different climate zones in India, providing key insights for enhancing drought prediction in India (Pathania and Gupta, 2025). Similarly, in a study by Ningchang G. et al., the model using Random Forest (RF) for feature selection and Extreme Gradient Boosting (XGBoost) enhanced by Bayesian Optimization achieved high accuracy in predicting greenhouse gas emission levels (Gan and Zhao, 2024). Subsequently, Bayesian networks have shown good adaptability in predicting the acute toxicity of chemicals to fish (Moe et al., 2020). Furthermore, Gaussian Process Regression (GPR) gives acceptable predictions of energy consumption in office buildings (Zeng et al., 2020). While machine learning methods possess certain adaptive capabilities, they are typically constrained in dealing with high-dimensional data and feature interactions.

In addition, with the rapid advancements in machine learning, deep learning models have emerged and have been employed at an exponential rate in energy forecasting. These models include convolutional neural networks (CNNs), temporal convolutional networks (TCNs), profile neural networks, and long short-term memory (LSTM) networks. In this context, the modified advanced convolutional neural network (CNN) U-Net can automatically identify river shapes and wet areas in flooded areas and achieve good results (Hosseiny, 2021). Inspired by graph learning, Zhen Shao et al. established a graph-based hybrid dynamic and static regional carbon emission network framework to perform real-time prediction of regional carbon emissions (Shao et al., 2024). Similarly, combining residual convolutional LSTM with attention mechanism achieves accurate spatiotemporal forest cover prediction (Liu et al., 2025). Deep learning methods have advantages in processing large-scale data. But when faced with data from different sources and multiple types of features, it may be challenging to fully identify and integrate the intricate associations between features. Also, there is a lack of effective feature selection and combination strategies. The essential characteristics of carbon emissions are complex and diverse. They involve factors such as natural gas consumption, coal consumption, and renewable energy generation, including solar and wind power. These factors are subject to varying degrees of interdependence with the target carbon emissions being predicted. The application of traditional statistical learning and machine learning is inherently constrained in its ability to address the complex and diverse characteristics. These limitations may result in the model's ability to fully extract information, thereby affecting the accuracy of the prediction. It is therefore imperative to develop new models and algorithms to help address the challenges of carbon emissions prediction.

The above factors serve as motivation for the presentation of our model, which is named as DEOH-XGBoost. The method is comprised of three principal components: feature engineering, model construction and model integration. In the initial phase, the process of feature engineering is utilized for the purpose of differentiating the multifarious correlations between the target indicator and an array of features. Features are grouped through membership function. Secondly, model construction involves creating four distinct deep learning models, each with a different feature set. Thirdly, model integration combines the four constructed deep learning models using a differential evolutionary weighting strategy. In our experiments, we collected real data on carbon emissions from seven regions (Africa, Asia and the Pacific, CIS, Europe, the Middle East, North America, and Central and South America) as targeted predictors. We also included 34 factors, such as natural gas consumption, coal consumption, and new energy generation, as features. This approach effectively reveals the intrinsic connections between different features and avoids including redundant or low-relevance features in the model. Our methodology has been validated through comprehensive experimentation on authentic datasets. The resultant findings demonstrate that the proposed approach exhibits superior accuracy in comparison to analogous state-of-the-art techniques.

2. Methodology

Fig. 1 illustrates the proposed model, which encompasses three stages: feature engineering, model construction, and model integration. The following section will provide a detailed explanation of the aforementioned three stages.

2.1. Feature engineering

Correlation is a measure of how closely two variables are linked. Correlation analysis studies the relationship between two or more variables. The aim is to determine whether there is a dependence between them. It explores the direction and degree of correlation for specific dependent phenomena. A crucial tool in correlation analysis is the calculation of correlation coefficients. These coefficients quantify the degree of correlation.

The correlation coefficient $COR(x, y)$ measures the correlation between x and y takes values in the range -1 to 1 . The stronger the correlation between the variables x and y , the closer the absolute value of $COR(x, y)$ is to 1 .

A variety of techniques for calculating correlation coefficients have been developed, including Spearman, Kendall, Pearson, partial correlation, and others. These methods employ different conceptual frameworks and measures of the relationship between the data. Spearman's method is based on the monotonic relationship between two variables. It uses the rank order of the variables as the basis for calculation. Furthermore, the consistency and inconsistency of the observed pairs is the basis for the Kendall correlation coefficient. Beside this, Pearson's correlation coefficient is a measure of the strength and direction of a linear relationship between two continuous variables. Finally, Partial correlation is a method that controls for the effect of one or more other variables. The application of these disparate correlation coefficient methodologies facilitates an enhanced understanding of the interrelationships between distinct types of variables, thereby enhancing the capacity to interpret and analyze data.

2.1.1. Pearson correlation analysis

The Pearson correlation coefficient measures the linear relationship between two variables. It ranges from -1 to 1 , with 1 being a perfect positive relationship, -1 being a perfect negative relationship and 0 being no relationship (Sedgwick, 2012). The definition is as follows:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

where X_i and Y_i are the data values of the feature X and the prediction target Y , respectively. \bar{X} and \bar{Y} are the means of X and Y , respectively.

2.1.2. Kendall rank correlation coefficient

The Kendall rank correlation coefficient is a non-parametric statistical measure used to assess the strength of the ordinal relationship between two variables (McLeod, 2005). The definition is as follows:

$$\tau = \frac{2}{n(n-1)} \sum_{i < j} \text{sign}(X_i - X_j) \text{sign}(Y_i - Y_j) \quad (2)$$

where X_i and Y_i are the data values of the feature and prediction respectively, and n is the number of data points.

2.1.3. Spearman's rank correlation coefficient

Spearman's rank correlation coefficient measures the monotonic relationship between two variables based on the ranks of the values (Xiao et al., 2016). The definition is as follows:

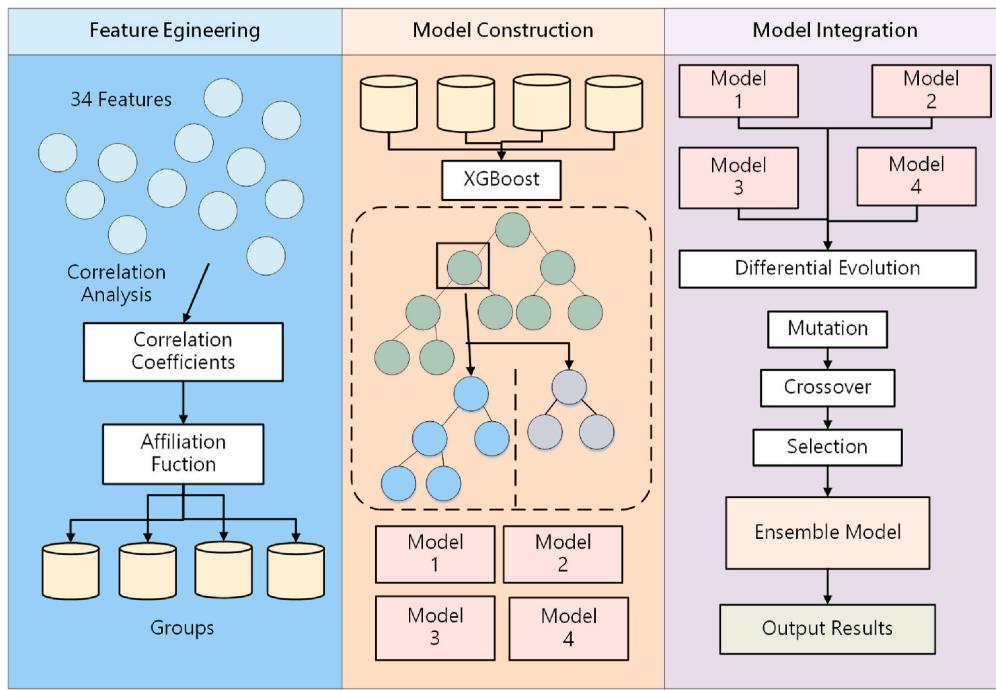


Fig. 1. Flowchart of proposed model.

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (3)$$

where d_i is the difference in ranks between the paired scores, here refers to the target and the feature. n is the number of data points.

2.1.4. Partial correlation coefficient

While controlling for the effect of one or more additional variables, the partial correlation coefficient measures the degree of association between two variables (Epskamp and Fried, 2018). The definition is as follows:

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (4)$$

where $r_{xy.z}$ is the partial correlation coefficient between X (the computed feature) and Y (the predicted target) controlling for Z , i.e., the other features; and r_{xz} and r_{yz} are the Pearson's correlation coefficients between X and Y , Z and Y , and Z and X .

2.1.5. Membership function

The concept of a membership function is of great significance in the fields of mathematics and fuzzy logic (Dombi, 1990). It is mainly used to represent the degree of membership of an element relative to a fuzzy set. Membership is from 0 to 1. 0 means that it is not part of this set. And a value of 1 indicates that the element is a full member of the set. Values between 0 and 1 represent the degree to which the element partially belongs to the set.

In the context of fuzzy set theory, traditional set theory employs clear boundaries to distinguish whether an element belongs to a set. These boundaries are often represented by binary logic, such as the statements 'true' or 'false', 'yes' or 'no'. Nevertheless, many phenomena and data set in the real-world lack clear boundaries and exhibit a certain degree of fuzziness. For example, temperature can be described as 'cold', 'warm' or 'hot' but there is no clear boundary between these descriptions. In such cases, the membership function is a valuable tool. It describes this fuzziness by defining a continuous membership degree.

One significant domain in which membership functions are utilized

is that of fuzzy decision support systems. In such systems, membership functions are employed to represent the extent to which different factors contribute to the decision target. In this experiment, membership functions are used to integrate features based on correlation coefficients. Features with similar correlations are grouped together.

Membership functions may assume a variety of forms, including triangular, trapezoidal, Gaussian, and bell-shaped functions, among others. The selection of these functions is contingent upon the particular application scenario. To illustrate, the triangular membership function (Azam et al., 2020) describes the change of membership through a linear relationship. This function is appropriate for some basic fuzzy sets. In contrast, the Gaussian membership function (Hameed, 2011) employs a normal distribution curve, which is more suitable for describing fuzzy sets with central symmetry.

In this experiment, a new idea was used to introduce a stepwise affiliation function based on the correlation coefficient which is shown on Fig. 2. The absolute value of the correlation coefficient, 0.5, is chosen as the starting point of the affiliation function, and 0.5 is assigned as the affiliation value at this starting point. When the absolute value of the correlation coefficient reaches 1, the corresponding degree of affiliation

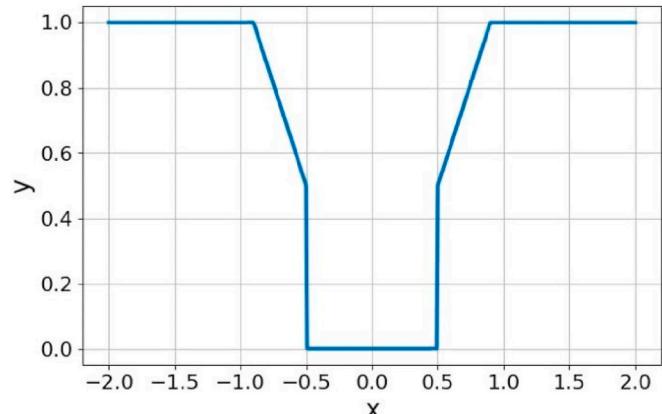


Fig. 2. Membership function.

is also 1, thus forming a trapezoidal curve and a mapping relationship that is symmetric about the y-axis. The four correlation coefficients are mapped onto this function to get the affiliation value. The features are selected in such a way that if the affiliation value of the feature under one of the correlation analyses is greater than the average of the four correlation analyses mapping affiliations, the feature is selected. The exact formula for calculating the affiliation value is given below.

$$f(x) = \begin{cases} 1, & \text{if } |x| \geq 0.9, \\ 0, & \text{if } |x| \leq 0.5, \\ 1.25x - 0.125, & \text{if } 0.5 < |x| < 0.9 \text{ and } x \geq 0, \\ -1.25x - 0.125, & \text{if } 0.5 < |x| < 0.9 \text{ and } x < 0. \end{cases} \quad (5)$$

where x refers to the correlation coefficient, and f refers to the affiliation value.

2.2. Model construction

XGBoost, or Extreme Gradient Boosting, is an advanced implementation of gradient boosting algorithms (see Fig. 3). The core principle behind XGBoost is the sequential addition of new models. These models correct the mistakes made by previous models. XGBoost utilizes a gradient boosting framework. To predict the residuals or errors of previous models, new models are created. These new models are then added together to make the final prediction (Hoffmann Souza et al., 2023).

First, we use the average carbon emission value as the initial prediction of the model $\hat{y}_i^{(0)}$. This provides a starting point for subsequent iterations.

For each sample i , we calculate the first-order derivative g_i and second-order derivative h_i of the loss function $l(y_i, \hat{y}_i)$ as follows:

$$g_i = \frac{\partial l(y_i, \hat{y}_i)}{\partial \hat{y}_i}, h_i = \frac{\partial^2 l(y_i, \hat{y}_i)}{\partial \hat{y}_i^2} \quad (6)$$

where g_i indicates how much adjustment is needed to the predicted carbon emissions \hat{y}_i to reduce the prediction error of sample i , and h_i indicates the rate of change of this adjustment.

We then iteratively build trees to continuously improve the model's prediction accuracy. For each new tree, we first find the best split point by calculating the gain at each split point as follows:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (7)$$

Where G_L and H_L are the sum of the first and second order gradients of the left node, and G_R and H_R are the sum of the right node. In DEOH-

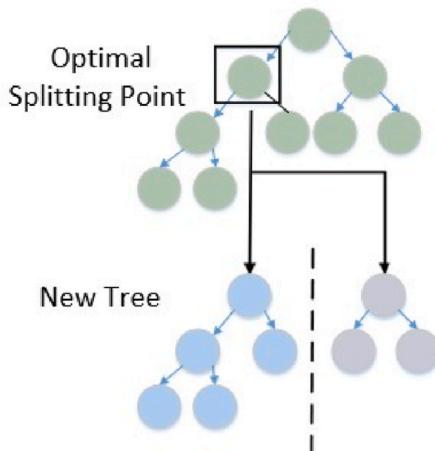


Fig. 3a. Tree construction process.

XGBoost, this gain represents the improvement in prediction accuracy obtained by splitting the data at this point.

After finding the best split point, we update the weights of the leaf nodes according to the best split point:

$$w_j = -\frac{G_j}{H_j + \lambda} \quad (8)$$

where w_j represents the updated predicted value of a specific leaf node, adjusting the predicted carbon emissions according to the gradient.

The new tree $f_t(x)$ we generated from the above steps is added to the model.

After each iteration, the output of the new tree is used to update the prediction of the overall model:

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_t(x_i) \quad (9)$$

where η is the learning rate, which controls the contribution of each tree to the final model. As we build the model, this update optimizes the predicted emissions by incorporating the latest adjustments from the new tree.

The final prediction model is the cumulative output of all trees:

$$\hat{y} = \sum_{k=1}^K f_k(x) \quad (10)$$

Hyperparameter tuning is a crucial step in optimizing the performance of XGBoost. In the course of tuning and testing the model with data from 1984 to 2022, several key hyperparameters were adjusted. These included the learning rate (η), maximum tree depth, subsample ratio, and regularization parameters (λ and γ). A grid search and stochastic search methodology (Cruz et al., 2021) were employed to investigate a range of potential hyperparameter values.

The objective of identifying the optimal configuration for the model. Combining these two methods effectively covers a wide range of parameter spaces. The generalizability and performance of the model is improved by this approach. Ultimately, it helps find the optimal hyperparameter settings, enhancing the accuracy and stability of machine learning models.

XGBoost has proven to be highly effective in predicting task due to its robustness and flexibility. By leveraging second-order gradient information, regularization, and various techniques to prevent overfitting, it achieves superior performance in predicting carbon emissions.

2.3. Model integration

Ensemble learning is a robust paradigm for machine learning. Multiple models are trained to solve the same problem, often referred to as 'base learners'. These models are then aggregated to produce a better outcome than any single model could achieve on its own. The central premise of ensemble methods is to combine the predictions of several base learners to enhance generalizability and robustness over a single learner (Dong et al., 2020). A number of methods have emerged, and here we employ differential evolution strategy to achieve model integration.

Differential Evolution (DE) is a stochastic and population-based algorithm for optimizing. It is designed for solving complex multi-modal optimization problems. DE is known for its simplicity, speed, and robustness. These qualities make it widely applicable across various scientific and engineering domains.

The mechanism of DE revolves around four main operations: initialization, mutation, crossover, and selection. These operations facilitate the exploration and exploitation of the search space to efficiently find the globally optimal solution (Mallipeddi et al., 2011). The process iterates over these operations until a stopping criterion is met. The following are the detailed steps, illustrated in Fig. 4.

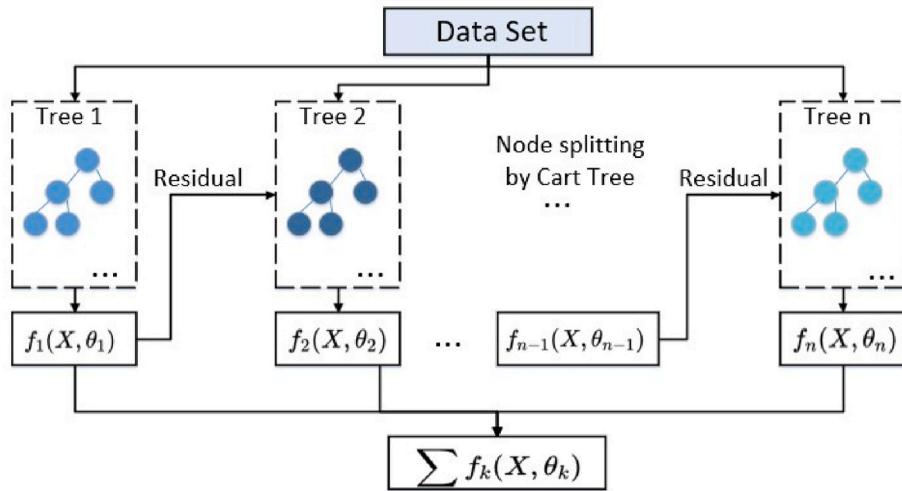


Fig. 3b. Algorithm of XGBoost.

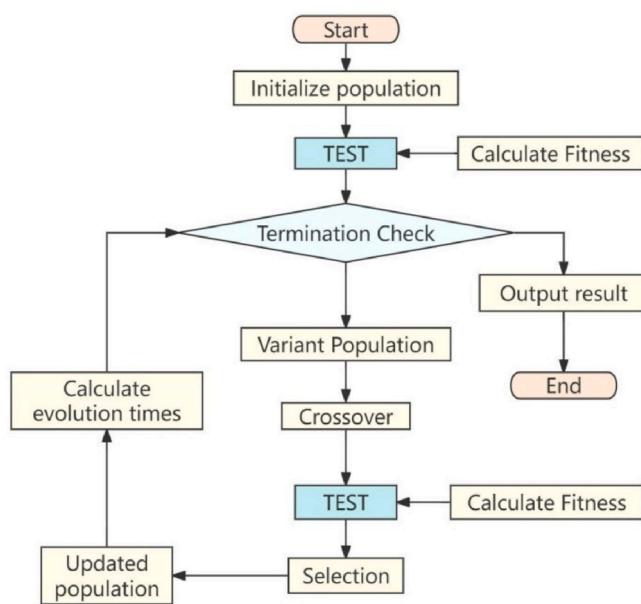


Fig. 4a. Algorithm of differential evolution.

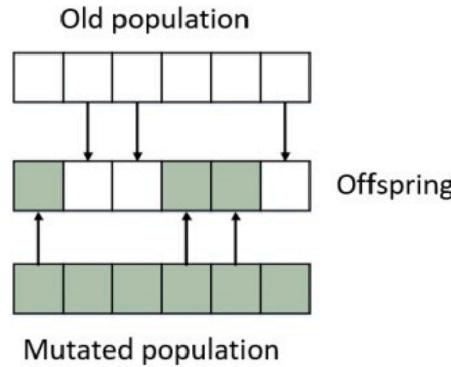


Fig. 4c. Offspring generation.

Step 1: Initialization

DE begins with the initialization of a population of potential solutions, termed as individuals. Each individual represents a potential solution to the optimization problem. The population size is generally constant throughout the process. The population is initialized uniformly at random within the defined parameter bounds. The target individual $I_{i,j}$ is defined as follows:

$$I_{i,j} = [I_{i,j}^1, I_{i,j}^2, \dots, I_{i,j}^\theta, \dots, I_{i,j}^3] \quad (11)$$

where $i \in \{1, 2, \dots, NP\}$ denotes the i -th individual, $j \in \{1, 2, \dots, \tau\}$ represents the j -th generation, and $\theta \in \{1, 2, 3, 4\}$ stands for the θ -th attribute of the individual. This attribute corresponds to the weights of four different XGBoost models trained on four different sets of features.

Step 2: Mutation

For each target individual $I_{i,j}$ in the current generation j , a mutant individual $O_{i,j}$ is generated. Our mutation strategy is DE/Rand/1, defined as:

$$O_{i,j} = I_{a1,j} + M_i \cdot (I_{a2,j} - I_{a3,j}) \quad (12)$$

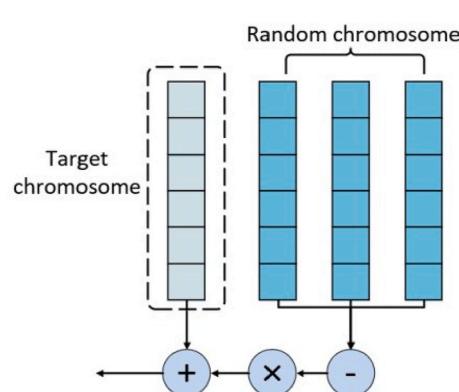


Fig. 4b. Mutation process.

where $I_{a1,j}$, $I_{a2,j}$, and $I_{a3,j}$, are different individuals randomly selected from the population. M_i is the scaling factor controlling the difference between two randomly selected individuals. The value of M_i is dynamically adjusted as follows:

$$M_i = \begin{cases} SFGSS & \text{if } \text{random}_3 < \delta_2 \\ SFHC & \text{if } \delta_2 \leq \text{random}_3 < \delta_3 \\ M_l + M_\mu \cdot \text{random}_1 & \text{if } \text{random}_2 < \delta_1 \text{ and } \text{random}_3 > \delta_3 \\ M_i & \text{otherwise} \end{cases} \quad (13)$$

where random1, random2, and random3 are pseudorandom numbers between 0 and 1. δ_1 , δ_2 , and δ_3 are constant threshold values.

Step 3: Crossover

The crossover operation combines the target individual $I_{i,j}$ and its mutant counterpart $O_{i,j}$ to produce a trial individual $F_{i,j}$. Our crossover strategy is arithmetic crossover:

$$F_{i,j} = I_{i,j} + \zeta \cdot (O_{i,j} - I_{i,j}) \quad (14)$$

where ζ is a random number from 0 to 1.

Step 4: Selection

After mutation and crossover, the selection operation determines which individuals survive to the next generation. The trial individual $F_{i,j}$ replaces the target individual $I_{i,j}$ if it has a better or equal objective function value OFV :

$$I_{i,j}^{(\text{new})} = \begin{cases} F_{i,j} & \text{if } OFV_{F_{i,j}} \leq OFV_{I_{i,j}} \\ I_{i,j} & \text{otherwise} \end{cases} \quad (15)$$

The objective function value (OFV) is calculated as the Mean Absolute Percentage Error (MAPE). This MAPE is derived from the combined predictions of the three XGBoost models. The predictions are weighted according to the weights represented by the individual. Consequently, the subsequent generation $j+1$ invariably outperforms or on par with the current generation j . Once the stopping criterion is met, an optimal population is achieved. The optimal individual F_{best} is selected to minimize the OFV in the population. Finally, the weights of the model are calculated using F_{best} .

We use Differential Evolution (DE) to achieve weight optimization and hence model integration to improve the accuracy of model predictions. Let y_i denote the prediction of the i -th model, w_i denotes the weight of the i -th model, n denotes the number of models, and y denotes the integrated model prediction. We aim to optimize the weights w_i using the Differential Evolution algorithm to minimize the prediction error. This process can be expressed as follows:

$$y = \sum_{i=1}^n w_i y_i \quad (16)$$

where w_i are the weights optimized using the DE algorithm.

To provide a more structured understanding, the following pseudo-code summarizes the DE procedure used for weight optimization:

Algorithm 1 Differential Evolution for Weight Optimization

```

1: Input: Initial population size  $NP$ , generation limit  $\tau$ , scaling factor  $M_b$ , crossover rate, objective function value ( $OFV$ ) based on MAPE.
2: Initialize: Randomly generate initial population  $I_{i,j}$ , where  $i \in \{1, \dots, NP\}$  and  $j = 1$ . Each individual contains four attributes representing the weights of four XGBoost models.

3: for generation  $j = 1$  to  $\tau$  do
4:   for each target individual  $I_{i,j}$  in population do
5:     /* Mutation */
6:     Randomly select three distinct individuals  $I_{a1,j}, I_{a2,j}, I_{a3,j} \neq I_{i,j}$ 
7:     Generate mutant vector  $O_{i,j} = I_{a1,j} + M * (I_{a2,j} - I_{a3,j})$  using formula 12
8:     Dynamically adjust  $M_i$  using formula 13 with  $\text{random1}, \text{random2}, \text{random3}$ 
9:     /* Crossover */
10:    Combine  $I_{i,j}$  and  $O_{i,j}$  using arithmetic crossover to produce  $F_{i,j}$  using formula 14
11:    /* Selection */
12:    Evaluate  $OFV(F_{i,j})$  and  $OFV(I_{i,j})$  using MAPE based on model predictions
13:    If  $OFV(F_{i,j}) \leq OFV(I_{i,j})$ , then replace:  $I_{i,j} \leftarrow F_{i,j}$ 
14:  end for
```

(continued on next column)

(continued)

Algorithm 1 Differential Evolution for Weight Optimization

```

15: end for
16: Identify best individual  $F_{best}$  with minimum  $OFV$  in final population
17: Compute integrated prediction  $y$  with formula 16 using weights from  $F_{best}$ 
18: Output: Optimized weights  $w_i$  and final ensemble prediction  $y$ 
```

2.4. Algorithm design

The algorithm for carbon emission projection is meticulously structured. It leverages a combination of data-driven techniques and advanced machine learning models.

Algorithm 2 Algorithm Design of Carbon Emission Projection

```

1: Input: Data on carbon emissions from various sources, divided into predictors and characteristics.
2: Initialize: Collect 35 factors from seven regions.

3: /* Feature Engineering */
4: Conduct correlation analysis using formula 1 to 4.
5: Select area features using formula 5.

6: /* Model Construction */
7: for each dataset  $D_i$  do
8:   Train XGBoost model  $M_i$  on  $D_i$  using formula 6 to 10.
9: end for
10: Obtain predictions  $y_i$  and errors  $e_i$ .

11: /* Model Integration */
12: for each generation  $g$  do
13:   Mutate using formula 12 and 13.
14:   Crossover using formula 14.
15:   Select using formula 15.
16: end for
17: Integrate models using formula 16.

18: Output: Projections and error metrics.
```

The process begins with collecting data on carbon emissions from various socio-economic and cultural sources across seven distinct regions. The data points, which include 35 factors, are categorized into two groups: targeted predictors (CO2 emissions from energy) and associated characteristics.

Next, a comprehensive feature engineering process is conducted. This involves performing correlation analyses, including Spearman's, Kendall's, Pearson's, and partial correlation. The objective is to derive correlation coefficients. Based on these coefficients, the area features are classified into three uniform groups using a specially proposed affiliation function mapping.

The core of the algorithm is the construction of a model using XGBoost (Extreme Gradient Boosting). XGBoost is chosen for its proven efficacy in similar predictive tasks. Three distinct datasets, each corresponding to a specific region, are used to train individual XGBoost models. These models are then fine-tuned through parameter optimization to achieve the most accurate predictions.

The integration of these models is achieved through a differential evolution (DE) approach. This approach optimizes the weights assigned to each model based on the mean absolute percentage error (MAPE) as a fitness measure. As a result, the integrated model provides a robust and accurate mechanism for projecting carbon emissions across the seven regions. The overall description of this algorithm is given in Algorithm 2.

3. Experiments and results

3.1. Data collection

The data utilized in this study was sourced from the 72nd Statistical Review of World Energy, published by the Energy Institute (EI) (energinst.org). The EI has been a chartered professional membership body

for over a century. It collaborates with industry partners and academic institutions. Together, they produce this comprehensive report. The Statistical Review is renowned for its detailed and reliable data on global energy production, consumption, trade, and emissions. The publication provides a yearly overview of the energy landscape, offering insights into the challenges posed by geopolitical and environmental crises. A total of 34 features were derived from multi-dimensional data for seven regions from 1985 to 2022. These included data on carbon emissions, carbon consumption, oil consumption and storage, natural gas consumption, and hydroelectricity generation, which is used in the next step of feature engineering. The features and calculation could be checked in [Table 1](#), and the regions are shown on [Fig. 5](#).

3.2. Experiment settings

3.2.1. Experiment design

In order to evaluate and demonstrate the effect achieved by DEOH-XGBoost, we designed the following experiment.

Experiment 1. Feature Engineering.

We analyzed the data from seven regions using Spearman, Pearson, Partial Correlation, and Kendall correlation analyses. Following this, the membership degree obtained will be used as the basis for selecting this feature. We have done this in each correlation analysis in each region.

Experiment 2. Comparative Experiment.

This experiment uses feature data from 1985 to 2020 and 1985 to 2021 to construct the training set, and uses indicator data from 2021 to 2022 to construct the test set for prediction, respectively. The results of the DEOH-XGBoost model and the experimental results of the single model are compared with the results of the currently popular single machine learning and deep learning models. This is done in order to verify that the proposed model is superior to the current prediction models. [Table 2](#) is an explanation of the single models and the DEOH-XGBoost mode.

Experiment 3. Ablation Study.

We designed an ablation study to validate the impact of different modules in the ensemble model on prediction accuracy. In this experiment, we systematically altered specific components of the model to observe the effect on overall performance ([Sime et al., 2023](#)).

We compare the results of these ablation tests with the full model. Thus, we were able to identify the critical components that significantly enhance the prediction accuracy and robustness of DEOH-XGBoost. This study provided valuable insights into the importance of each component and helped to finetune the model for better performance.

Experiment 4. Hyperparameter Analysis.

The objective of this experiment was to conduct a comprehensive hyperparameter sensitivity analysis of the differential evolution model. The analysis aimed to elucidate the pivotal role that these hyperparameters play in the model's functioning. The experiment

Table 1

Data Collection Status. We collected data on carbon emission indicators from 1985 to 2022 for seven countries, with a total of 9310 records.

Year	Regions	Features (excepting indicators)
1985 to 2022	Africa, Asia and Pacific, CIS, Europe, Middle East, North America, South and Central America	Coal Consumption, Primary Energy Consumption, Nuclear Consumption, Oil Consumption, Natural Gas Consumption, Oil Production, Oil Refining Capacity, Natural Gas reserves, Coal Production, Nuclear Generation, Hydroelectricity Consumption, Renewables Generation, Wind Generation, Wind consumption, Solar Generation, Solar consumption, Hydroelectricity Generation, Oil consumption by product, Primary Energy Consumption, Electricity From Coal, Electricity From Gas, Electricity From Oil, Electricity From Other, Electricity Generation, Gas Production, Oil Proved Reserves, etc.
	Total of 7	Total of 245

investigated the impact of different hyperparameter combinations. It demonstrated how alterations in hyperparameters influence the performance and ultimate outcome of the model.

3.2.2. Evaluation metric

- MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error) ([De et al., 2016](#)) is used to measure the error between the prediction result and the real value, which is described as follows:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{A_i - F_i}{A_i} \right| \times 100\% \quad (17)$$

where A_i represents the actual values of carbon emission, F_i stands for the predicted values, and n is the total number of data points in the dataset.

- RMSE (Root Mean Squared Error)

RMSE (Root Mean Squared Error) is used to measure the deviation between the predicted values and the actual values ([Fisher, 1920](#)), which is defined as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (A_i - F_i)^2} \quad (18)$$

where A_i represents the actual values of carbon emission, F_i stands for the predicted values, and N is the total number of data points in the dataset.

3.3. Feature engineering

In order to systematically group and filter the features, we performed four separate correlation analyses on the features. To illustrate, the features of the African region are shown in [Tables 3–6](#). It can be seen that the features of the African region are generally strongly correlated with the predicted object. Subsequently, we mapped the features using the designed correlation function to get the affiliation of each feature. The features were selected in such way that if the affiliation of the feature under this correlation analysis was greater than the average of the affiliations obtained from the mapping of the four correlation analyses, it was selected. This process generated four different sets of features for each region, corresponding to four different correlation analyses, which were then used for constructing the four models.

3.4. Comparative experiment

The present set of experiments aims to establish a comparison between the integrated model that has been proposed and five other individual models. The five models in question are as follows: LSTM ([Li et al., 2023](#)), TCN, CNN ([Xiao et al., 2022](#)), TRMF and XGBoost ([Hou et al., 2023](#)). The experimental results are displayed in [Tables 7 and 8](#).

In order to facilitate a more profound comprehension of the outcomes of these comparisons, statistical analyses were conducted utilizing the findings of preceding studies on loss/win, Wilcoxon signed rank test ([Woolson, 2007](#)) and Friedman test ([Zimmerman and Zumbo, 1993](#)).

The table presents the performance of various models, including CNN, TCN, LSTM, XGBoost, TRMF, and our DEOH-XGBoost. It covers different regions: North America, Central and South America, Europe, CIS, Middle East, and Africa. Performance is evaluated through measurement of error rates. It is widely accepted that lower values are indicative of improved performance.

The objective of this comparative analysis is to elucidate the merits

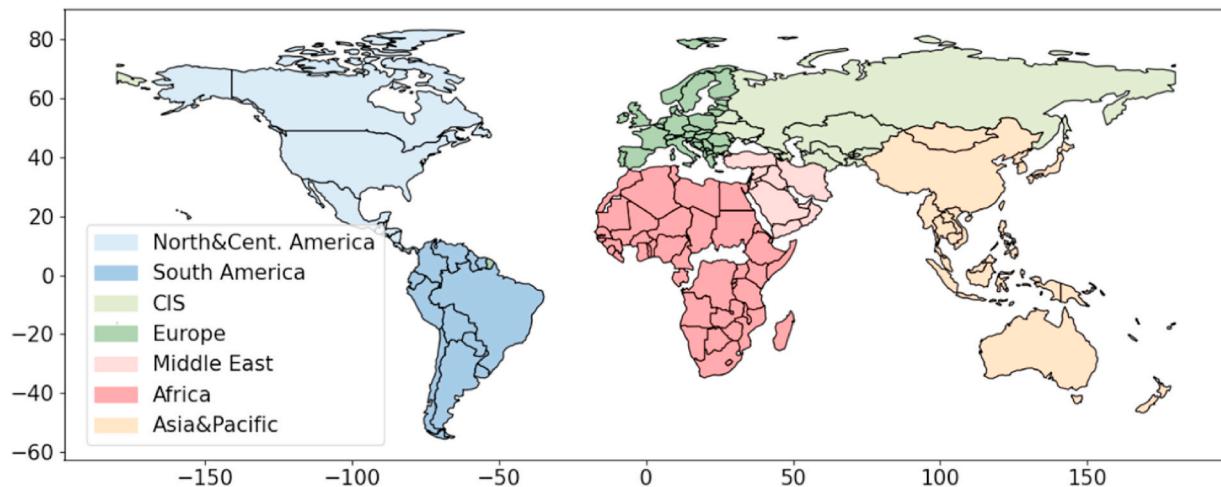


Fig. 5. World regions corresponding to predictive indicators.

Table 2

Description of the individual models and the DEOH-XGBoost model. This includes CNN, TCN, LSTM, XGBoost, TRMF and our model DEOH-XGBoost.

Model	Description
CNN	CNN (Convolutional Neural Network) uses convolutional layers to learn spatial hierarchies of features automatically and adaptively.
TCN	TCN (Temporal Convolutional Network) uses 1D fully-convolutional networks with causal convolutions and dilated convolutions to capture long-range dependencies in sequential data.
LSTM	LSTM (Long Short-Term Memory) addresses the vanishing gradient problem by using memory cells to maintain information over extended periods.
XGBoost	XGBoost provides parallel tree boosting to solve classification and regression problems, known for its speed and performance.
TRMF	TRMF (Temporal Regularized Matrix Factorization) combines matrix factorization with temporal regularization to capture the temporal dynamics in time series data.
DEOH-XGBoost	The model employed in this study based on the XGBoost algorithm and incorporates feature engineering, differential evolution, and model integration.

Table 3

The results of the Spearman correlation coefficients of Africa's carbon emission indicators. The correlation coefficient ranges are 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0., respectively.

Range	Feature
0–0.2	Carbon Dioxide Emissions from Flaring, Natural Gas Flaring, Oil Refinery throughput
0.2–0.4	Nuclear Consumption
0.4–0.6	Primary Energy Consumption per capita Gigajoule per capita
0.6–0.8	Oil Production, Nuclear Generation
0.8–1	Coal Consumption, Primary Energy Consumption, Oil Consumption, Natural Gas Consumption, Oil Refining Capacity, Oil Total liquids consumption in thousands of barrels per day, Natural Gas Proved reserves, Coal Production, Hydroelectricity Consumption, Renewables Generation Exajoules, Renewables Geothermal Biomass Other Exajoules, Renewables Geothermal Biomass Other TWh, Renewables Wind Generation TWh, Renewables Wind consumption Exajoules, Renewables Solar Generation TWh, Renewables Solar consumption Exajoules, Hydroelectricity Generation TWh, Oil Regional consumption by product Barrels, Primary Energy Consumption Million tonnes oil equivalent, Electricity From Coal, Electricity From Gas, Electricity From Oil, Electricity From Other, Electricity Generation, Gas Production, Oil Proved Reserves

and limitations of each model. It particularly focuses on the advantages of the Ensemble Model.

The experimental results for 2021 and 2022 demonstrate the superiority of the DEOH-XGBoost model over other comparative models in

Table 4

The results of the Pearson correlation coefficients of Africa's carbon emission indicators. The correlation coefficient ranges are 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0., respectively.

Range	Feature
0–0.2	Oil Refinery throughput
0.2–0.4	Carbon Dioxide Emissions from Flaring, Natural Gas Flaring
0.4–0.6	Nuclear Consumption, Oil Production, Primary Energy Consumption per capita Gigajoule per capita
0.6–0.8	Nuclear Generation, Renewables Wind Generation TWh, Renewables Wind consumption Exajoules, Renewables Solar Generation TWh, Renewables Solar consumption Exajoules
0.8–1	Coal Consumption, Primary Energy Consumption, Oil Consumption, Natural Gas Consumption, Oil Refining Capacity, Oil Total liquids consumption in thousands of barrels per day, Natural Gas Proved reserves, Coal Production, Hydroelectricity Consumption, Renewables Generation Exajoules, Renewables Geothermal Biomass Other Exajoules, Renewables Geothermal Biomass Other TWh, Hydroelectricity Generation TWh, Oil Regional consumption by product Barrels, Primary Energy Consumption Million tonnes oil equivalent, Electricity From Coal, Electricity From Gas, Electricity From Oil, Electricity From Other, Electricity Generation, Gas Production, Oil Proved Reserves

Table 5

The results of the partial correlation coefficients of Africa's carbon emission indicators. The correlation coefficient ranges are 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0., respectively.

Range	Feature
0–0.2	Carbon Dioxide Emissions from Flaring, Natural Gas Flaring
0.2–0.4	Oil Refinery throughput
0.4–0.6	Renewables Solar Generation TWh, Renewables Solar consumption Exajoules
0.6–0.8	Oil Production, Renewables Generation Exajoules, Renewables Wind Generation TWh, Renewables Wind consumption Exajoules
0.8–1	Coal Consumption, Primary Energy Consumption, Nuclear Consumption, Oil Consumption, Natural Gas Consumption, Oil Refining Capacity, Oil Total liquids consumption in thousands of barrels per day, Natural Gas Proved reserves, Coal Production, Nuclear Generation, Hydroelectricity Consumption, Renewables Geothermal Biomass Other Exajoules, Renewables Geothermal Biomass Other TWh, Hydroelectricity Generation TWh, Oil Regional consumption by product Barrels, Primary Energy Consumption per capita Gigajoule per capita, Electricity From Coal, Electricity From Gas, Electricity From Oil, Electricity From Other, Electricity Generation, Gas Production, Oil Proved Reserves

Table 6

The results of the Kendall's rank correlation coefficients of Africa's carbon emission indicators. The correlation coefficient ranges are 0–0.2, 0.2–0.4, 0.4–0.6, 0.6–0.8, and 0.8–1.0, respectively.

Range	Category
0–0.2	Oil Refinery throughput
0.2–0.4	Carbon Dioxide Emissions from Flaring, Natural Gas Flaring, Nuclear Consumption, Primary Energy Consumption per capita Gigajoule per capita
0.4–0.6	Oil Production, Nuclear Generation
0.6–0.8	Oil Refining Capacity, Electricity From Coal, Electricity From Oil, Electricity From Other
0.8–1	Coal Consumption, Primary Energy Consumption, Oil Consumption, Natural Gas Consumption, Oil Total liquids consumption in thousands of barrels per day, Natural Gas Proved reserves, Coal Production, Hydroelectricity Consumption, Renewables Generation Exajoules, Renewables Geothermal Biomass Other Exajoules, Renewables Geothermal Biomass Other TWh, Renewables Wind Generation TWh, Renewables Wind consumption Exajoules, Renewables Solar Generation TWh, Renewables Solar consumption Exajoules, Hydroelectricity Generation TWh, Oil Regional consumption by product Barrels, Primary Energy Consumption Million tonnes oil equivalent, Electricity From Gas, Electricity Generation, Gas Production, Oil Proved Reserves

terms of both MAPE and RMSE. The key observations are summarized as follows:

- DEOH-XGBoost achieves the lowest mean MAPE and RMSE across both years, significantly outperforming CNN, TCN, LSTM, XGBoost, and TRMF. The consistently low errors confirm its effectiveness in time-series forecasting.
- The Wilcoxon signed-rank test results indicate that the performance improvement of DEOH-XGBoost is statistically significant, with all p-values below 0.05 when compared to alternative models. Furthermore, its F-rank remains the lowest (1.500), demonstrating its consistent accuracy.
- Unlike other models that exhibit fluctuating performances across different regions, DEOH-XGBoost maintains high accuracy across all experimental settings. This highlights its strong generalization ability and stability in various forecasting scenarios.
- DEOH-XGBoost achieves the best Win/Loss ratio (8/72) compared to other models, reinforcing its dominance in predictive performance.

Table 7

Results are presented for the MAPE and RMSE comparison between the single models and the proposed integrated approach for the 2021 forecast for seven regions. This comparison demonstrates that the proposed approach significantly outperforms the five single comparison models.

Regions	Deviation	CNN	TCN	LSTM	XGBoost	TRMF	DEOH-XGBoost
Africa	MAPE	3.805 %	2.714 %	2.730 %	5.437 %	3.551 %	0.831 %
	RMSE	49.709	35.460	35.669	71.025	46.389	10.850
Asia	MAPE	3.482 %	4.478 %	4.831 %	3.660 %	6.189 %	3.458 %
	RMSE	615.691	791.699	854.219	647.129	1094.344	611.366
CIS	MAPE	2.810 %	1.473 %	5.065 %	3.791 %	3.733 %	2.226 %
	RMSE	60.188	31.550	108.490	81.211	79.953	47.689
Europe	MAPE	4.566 %	5.537 %	1.554 %	1.731 %	9.417 %	1.060 %
	RMSE	175.566	212.902	59.753	66.546	362.094	40.763
Middle East	MAPE	4.535 %	5.207 %	2.752 %	3.791 %	9.626 %	3.717 %
	RMSE	95.291	109.406	57.835	81.211	202.263	78.108
S&Central America	MAPE	5.939 %	0.705 %	5.000 %	5.599 %	4.913 %	0.242 %
	RMSE	74.093	8.792	62.376	117.647	61.292	3.024
North America	MAPE	4.756 %	4.243 %	0.192 %	4.669 %	4.830 %	4.301 %
	RMSE	272.423	243.073	10.997	267.482	276.683	246.394
Mean *	MAPE	4.270 %	3.479 %	3.161 %	4.097 %	6.037 %	2.262 %
	RMSE	191.852	204.698	169.906	190.322	303.288	148.313
Loss/Win •		0/16	4/12	4/12	0/16	0/16	8/72
F-rank *		4.125	3.125	3.000	4.125	5.125	1.500
p-value ◊		0.00024	0.01069	0.04643	0.00024	0.00024	—

*Mean MAPE and RMSE of the seven regions.

• The present approach does not facilitate comparison, i.e. the row of 'MAPE' in 'CIS' and the column of 'TCN'.

◊ A lower value is indicative of enhanced forecasting accuracy.

◊ Our approach demonstrates a significantly higher degree of efficacy than the comparison model when the p-value was found to be lower than 0.05.

- While other models show variations in error rates, DEOH-XGBoost consistently produces the lowest RMSE and MAPE values. This suggests that the model effectively captures complex temporal dependencies and mitigates forecasting errors better than traditional methods.

Overall, DEOH-XGBoost shows excellent prediction accuracy, stability, and robustness across different regions and years. This highlights that our proposed model has lower prediction errors and higher prediction accuracy. It provides more accurate and reliable results for carbon emission forecasts.

3.5. Cross-regional feature selection comparison

To investigate the consistency and heterogeneity of selected features across diverse regional contexts, we conducted a cross-regional comparison using Jaccard similarity coefficients and UpSet plot visualization (Figs. 6 and 7). The seven aggregated regions include Africa (Af), Asia (As), CIS, Europe (Eu), Middle East (Me), South and Central America (Sc), and Northern America (Nm).

Jaccard Similarity Insights

As shown in Fig. 6, the Jaccard similarity heatmap reveals several clear groupings and divergences:

- Africa and Asia exhibit the highest mutual similarity, indicating that their carbon emission patterns are driven by similar underlying factors.
- South & Central America also shows high alignment with Asia and Africa, suggesting shared energy and economic characteristics among developing and emerging economies.
- In contrast, Northern America has notably low similarity with other regions, indicating a distinct set of emission drivers.
- The CIS region sits in an intermediate position, with moderate overlap with Europe but low similarity with other developing regions.

UpSet Plot Interpretation.

The UpSet plot in Fig. 7 further illustrates the intersectional structure of feature selection:

Table 8

Results are presented for the MAPE and RMSE comparison between the single models and the proposed integrated approach for the 2022 forecast for seven regions. This comparison demonstrates that the proposed approach significantly outperforms the five single comparison models.

Regions	Deviation	CNN	TCN	LSTM	XGBoost	TRMF	DEOH-XGBoost
Africa	MAPE	0.439 %	4.813 %	5.130 %	1.886 %	6.304 %	0.118 %
	RMSE	5.735	62.898	67.000	24.645	82.371	1.536
Asia	MAPE	5.373 %	2.892 %	5.398 %	2.039 %	3.228 %	2.013 %
	RMSE	964.815	519.260	969.158	366.096	579.655	361.521
CIS	MAPE	3.233 %	3.021 %	3.170 %	3.356 %	13.629 %	0.970 %
	RMSE	65.734	61.424	64.454	68.232	277.067	19.715
Europe	MAPE	3.082 %	5.664 %	2.759 %	0.579 %	6.988 %	0.251 %
	RMSE	116.200	213.539	104.007	21.822	263.427	9.481
Middle East	MAPE	6.828 %	2.558 %	8.221 %	4.453 %	6.564 %	4.429 %
	RMSE	150.204	56.284	180.868	97.967	144.398	97.433
S&Central America	MAPE	1.915 %	0.997 %	2.383 %	2.090 %	5.909 %	0.893 %
	RMSE	24.084	12.540	29.978	26.285	74.336	11.233
North America	MAPE	0.609 %	5.234 %	1.089 %	3.309 %	2.349 %	2.970 %
	RMSE	35.610	306.287	63.732	193.623	137.439	173.800
Mean *	MAPE	3.068 %	3.597 %	4.021 %	2.530 %	6.424 %	1.663 %
	RMSE	194.626	176.033	211.314	114.096	222.670	96.388
Loss/Win •		2/14	2/14	2/14	0/16	2/14	8/72
F-rank *		3.438	3.313	4.375	3.250	5.125	1.500
p-value ◊		0.00702	0.00283	0.00387	0.00024	0.00147	—

* Mean MAPE and RMSE of the seven regions.

• The present approach does not facilitate comparison, i.e., the row of ‘MAPE’ in ‘North America’ and the column of ‘CNN’.

* A lower value is indicative of enhanced forecasting accuracy.

◊ Our approach demonstrates a significantly higher degree of efficacy than the comparison model when the p-value was found to be lower than 0.05.

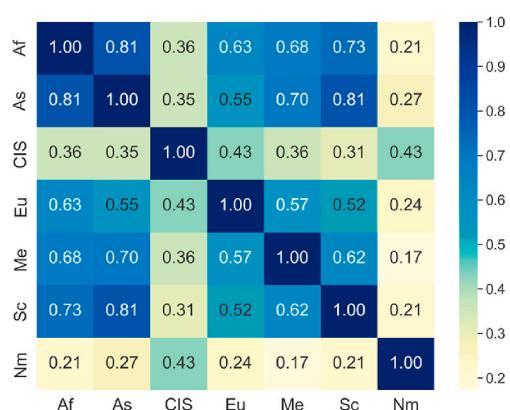


Fig. 6. Jaccard similarity across regions.

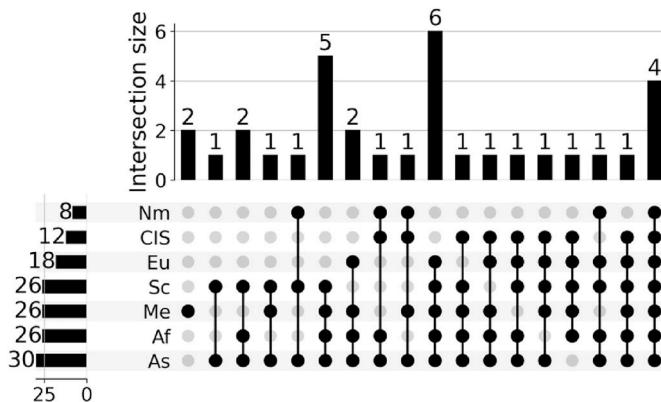


Fig. 7. Feature overlap across regions.

• Only a limited number of features are shared across a majority of regions, indicating strong regional differentiation.

• Africa, Asia, Middle East, and South America form the largest intersection, supporting the earlier Jaccard-based grouping.

• Many features are region-specific, with frequent 1-feature intersections, confirming the contextual uniqueness of emission determinants.

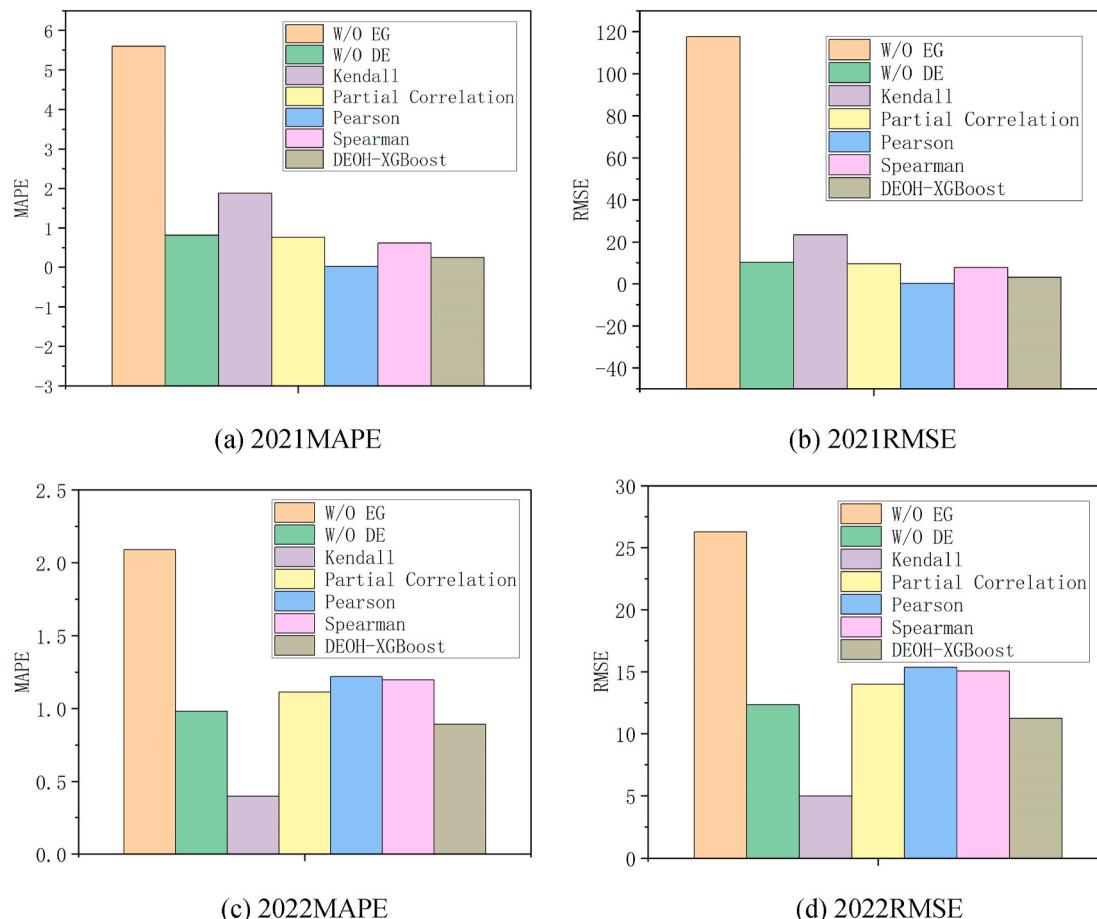
• Europe and North America contribute relatively fewer overlapping features, reinforcing the observation that climate-mature or transition economies require tailored modeling.

Collectively, these findings emphasize the substantial regional heterogeneity in carbon emission drivers and feature relevance. The integration of Jaccard similarity analysis and intersection-based visualization techniques has been demonstrated to substantiate the necessity for emission prediction models to encompass both shared global patterns and pronounced local distinctions. The model under discussion has been developed to capture this complexity by adapting feature selection to regional contexts, thereby enabling more accurate and context-sensitive forecasting across diverse economic and environmental landscapes.

3.6. Ablation study

To verify the impact of different modules in the ensemble model on forecast accuracy, we performed an ablation study using forecasts for 2022 and 2021. Taking the region South and Central America as an example, Fig. 8 below illustrates the mean absolute percentage error (MAPE) and root mean square error (RMSE) of different ensemble models.

In Fig. 8, ‘W/O EG’ indicates the experiment without grouping of features. ‘Kendall’, ‘Partial Correlation’, ‘Pearson’ and ‘Spearman’ refer to the four models based on different features. ‘W/O DE’ indicates that differential evolutionary integration is not applied, and the mean integration strategy is used instead. Beside this, ‘DEOH-XGBoost’ is the model we proposed. The figures allow us to derive the following conclusions:

**Fig. 8.** Ablation study.

- The removal of the feature grouping mechanism significantly increases both MAPE and RMSE. This result indicates that directly using all features without structured categorization can lead to redundant information and an increase in noise, making it more difficult for the model to learn meaningful patterns. Feature grouping helps the model focus on the most informative aspects of the data, reducing overfitting and improving generalization. The large error observed in this configuration demonstrates that grouping features based on their relationships enhances the forecasting ability by structuring the input space more effectively.
- When the differential evolutionary (DE) integration is removed and replaced with a simple mean integration strategy, a notable increase in forecasting errors is observed. The DE integration plays a crucial role in adjusting the weight distributions of different base models, thereby ensuring that the most reliable predictions contribute more significantly to the final output. In contrast, a simple mean strategy assigns equal importance to all models, disregarding their individual forecasting accuracy. The increased error in this setting highlights that an optimization-based integration approach is necessary to fully exploit the strengths of different ensemble components.
- The four models employing different statistical correlation techniques exhibit varying levels of forecasting accuracy. These methods are used to select features based on their relationships with the target variable, impacting the overall model performance. The results confirm that appropriate feature selection is critical to improving forecasting accuracy.
- The proposed DEOH-XGBoost model achieves the lowest MAPE and RMSE among all tested configurations. This demonstrates that the combination of feature grouping, correlation-based feature selection, and differential evolutionary optimization effectively enhances

forecasting accuracy. The consistently superior performance of DEOH-XGBoost across multiple datasets further validates its robustness in time-series forecasting applications.

The importance of feature engineering, model ensemble, and optimization techniques in building robust predictive models is shown by the results. The excellent performance of ensemble model validates the effectiveness of the enhancements proposed in the ensemble learning framework.

3.7. Hyperparameter Analysis

The present study conducts a hyperparameter sensitivity analysis with the objective of investigating the impact of four critical parameters on the performance of the DE model: Mutation Factor (F), Crossover Probability (CR) and Number of Particles, which could be seen on Fig. 9.

From the experiment results, the following statements could be concluded:

- It is observed that when CR is set to zero, both MAPE and RMSE are significantly high. As CR increases, the error metrics rapidly decrease and then stabilize. This suggests that a higher crossover probability facilitates better exploration of the search space, leading to improved optimization results.
- The model achieves relatively low MAPE and RMSE values when $F \geq 0.2$, while values below this threshold lead to noticeable performance degradation. The results indicate that selecting an appropriate mutation factor is crucial for balancing exploration and exploitation in the DE algorithm.

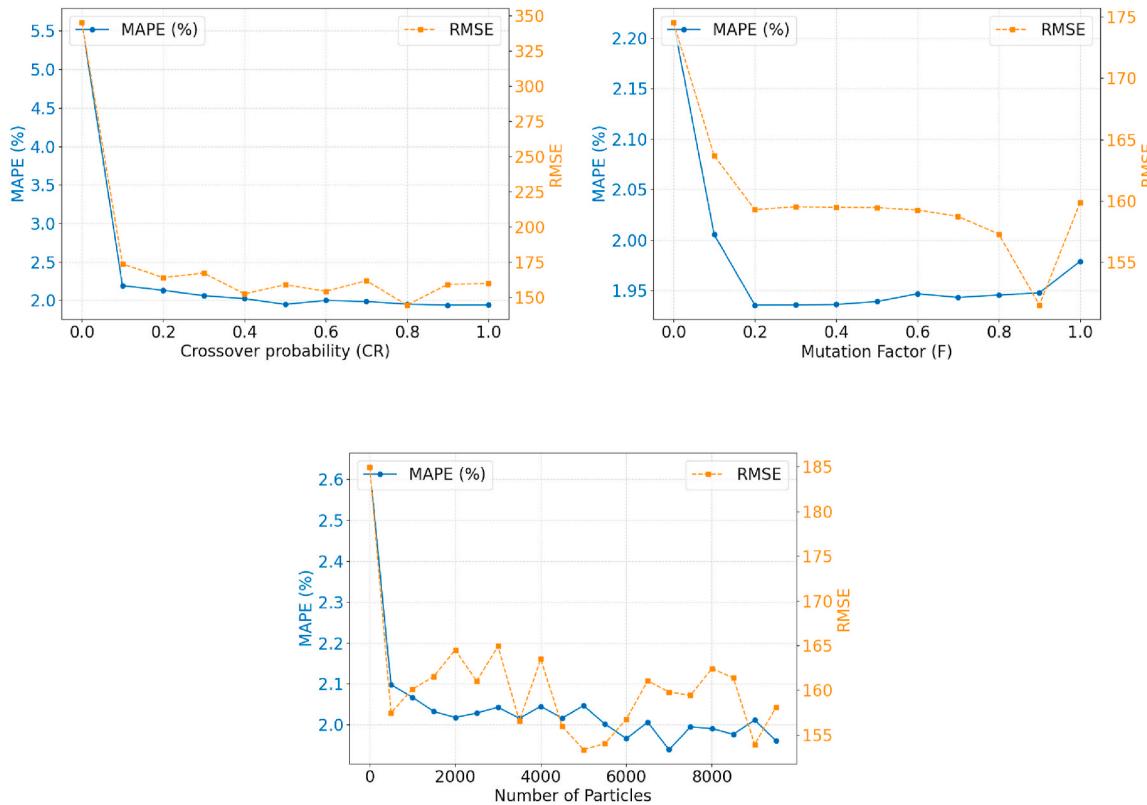


Fig. 9. Hyperparameter analysis.

- The results demonstrate that increasing the number of particles significantly reduces both MAPE and RMSE at the beginning. However, beyond a certain point, the improvement plateaus, suggesting that excessive particles do not necessarily enhance performance. This highlights the importance of selecting an optimal particle size to balance computational efficiency and prediction accuracy.

In conclusion, hyperparameter tuning is essential for optimizing the performance of the Differential Evolution (DE) model, as improper parameter selection can lead to suboptimal results with increased prediction errors and reduced efficiency. The analysis demonstrates that an appropriate crossover probability enhances solution exploration and prevents premature convergence, while a well-chosen mutation factor balances exploration and exploitation, ensuring stable optimization. Additionally, the number of particles significantly influences convergence speed and computational cost, with diminishing returns beyond a certain threshold. These findings highlight the necessity of systematic hyperparameter tuning to achieve robust model performance.

4. Case study case-study

In order to validate the robustness and adaptability of the proposed DEOH-XGBoost model, a case study was conducted across four representative countries. The following countries are involved: China, the United States, the United Kingdom and France. The countries in question differ significantly in terms of their energy structures, economic development stages, and carbon emission patterns, rendering them ideal subjects for evaluating the model's generalizability and flexibility.

4.1. Model performance comparison

To quantitatively evaluate the predictive capability of the proposed DEOH-XGBoost model, we conducted comparative experiments against several established baseline models: CNN, TCN, standard XGBoost,

TRMF, and LSTM. The evaluation metrics included Mean Absolute Percentage Error (MAPE) and Root Mean Squared Error (RMSE), measured across four representative countries: China, the United States, the United Kingdom, and France.

As shown in Table 9, the DEOH-XGBoost model consistently outperforms all other models in terms of both MAPE and RMSE across the majority of countries. Specifically:

- In all four evaluated countries, DEOH-XGBoost achieves the lowest MAPE and RMSE, demonstrating superior accuracy and robustness compared to conventional machine learning and deep learning models including CNN, TCN, XGBoost, TRMF, and LSTM. This consistent superiority across geographically and structurally diverse regions suggests the model's strong generalization ability.
- On average, DEOH-XGBoost reduces the mean MAPE to 1.96 and RMSE to 140.59, far outperforming the best-performing baseline. The model records a perfect Loss/Win ratio of 0/50, meaning it ranks first in all 50 comparisons across countries and metrics. Statistical analysis using the Friedman test yields a p-value of 0.019, confirming that the performance gains are statistically significant. Furthermore, the model obtains the lowest F-rank, reflecting its consistent top performance.

These results validate the effectiveness of the DEOH-XGBoost model for carbon emission prediction, particularly in capturing regional heterogeneity and minimizing forecast errors in complex multi-source datasets.

4.2. Statistical similarity analysis

One of the key strengths of our model lies in its adaptive feature selection mechanism, which combines multiple correlation metrics and fuzzy membership evaluation to extract the most influential drivers of emissions for each country. To further explore the diversity and overlap

Table 9

Results are presented for the MAPE and RMSE comparison between the single models and the proposed integrated approach for the 2022 forecast for four countries. This comparison demonstrate that the proposed approach significantly outperforms the five single comparison models.

Countries	Deviation	CNN	TCN	XGBoost	TRMF	LSTM	DEOH-XGBoost
China	MAPE	4.56	9.07	5.67	10.33	7.25	3.92
	RMSE	480.91	957.16	598.46	1090.15	765.20	413.34
US	MAPE	5.56	4.73	3.38	3.51	4.91	3.03
	RMSE	268.32	228.45	163.32	169.36	237.14	146.23
United Kingdom	MAPE	5.27	3.82	3.82	3.94	2.04	0.51
	RMSE	18.15	13.15	13.16	13.56	7.04	1.74
France	MAPE	1.83	8.71	6.71	3.90	2.28	0.38
	RMSE	4.94	23.48	18.09	10.51	6.15	1.03
Mean	MAPE	4.30	6.58	4.90	5.42	4.12	1.96
	RMSE	193.08	305.56	198.26	320.90	253.88	140.59
Loss/Win •		0/10	0/10	0/10	0/10	0/10	0/50
F-rank *		3.600	4.650	3.350	4.600	3.300	1.500
p-value ◇		0.019	0.019	0.019	0.019	0.019	—

* Mean MAPE and RMSE of the seven regions.

• The present approach does not facilitate comparison, i.e. the row of 'MAPE' in 'CIS' and the column of 'TCN'.

*A lower value is indicative of enhanced forecasting accuracy.

◇ Our approach demonstrates a significantly higher degree of efficacy than the comparison model when the p-value was found to be lower than 0.05.

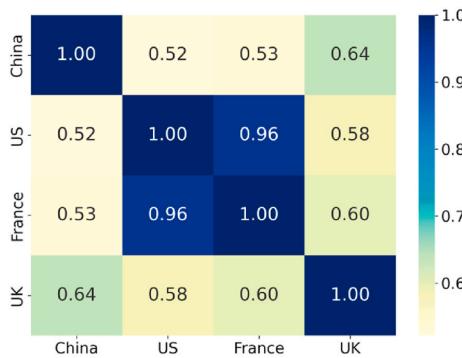


Fig. 10. Jaccard similarity across countries.

of selected features across the four representative countries, we conducted a quantitative analysis using Jaccard similarity coefficients and UpSet plot visualization.

As illustrated in Fig. 10, the pairwise similarity between national feature sets reveals varying degrees of overlap, with some notable regional patterns:

- The United Kingdom and China exhibit high Jaccard similarity at 0.64, indicating a relatively strong overlap in selected features.
- The United States and France also demonstrate a high degree of similarity, suggesting that their carbon emission patterns may be influenced by comparable drivers.
- France and the United Kingdom share a similarity score of 0.60, while US and UK are moderately aligned at 0.58, implying partial convergence in feature importance, possibly due to shared policy frameworks or economic structures.
- In contrast, China vs. US and China vs. France exhibit relatively low similarity, highlighting divergent feature selection patterns.

In order to facilitate a more intuitive visualization of the commonality and uniqueness of the selected features, an UpSet plot was employed in Fig. 11. The plot demonstrates that:

- 21 features are commonly shared by all four countries (China, UK, US, and France), suggesting the presence of a globally consistent core set of predictive variables, likely tied to universally important factors such as energy production, industrial output, and CO₂ emissions.
- China has the largest individual feature set, followed by the UK, US, and France, indicating regional variability in feature importance.

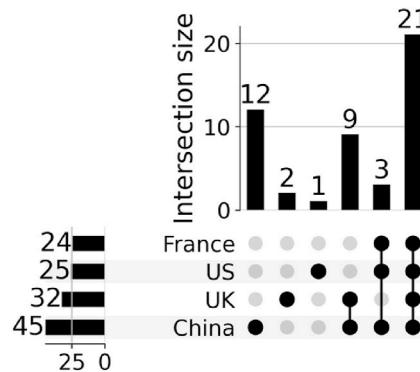


Fig. 11. Feature overlap across countries.

- France and China share 12 features exclusively, while other combinations—such as the UK, US, and China—only share 3 features, and UK and US alone share just 1 feature.

In summary, while a globally shared subset of features exists across all four countries, the observed variations in pairwise overlaps and country-specific selections reflect the diverse emission contexts and data characteristics. The findings demonstrate that the model does not rely on a one-size-fits-all approach, but instead adapts its feature selection to the specific conditions of each country. This adaptability is essential for capturing regional emission dynamics and enhancing the robustness and generalizability of the predictive framework.

4.3. Insights and implications

The results of this study provide both methodological and environmental insights. The varying feature selections and predictive performances across countries reflect the divergent structural, economic, and environmental conditions shaping national carbon emission dynamics.

In China, the dominant features relate to coal consumption, production, and industrial energy use, highlighting the country's continued reliance on coal and energy-intensive industries. These patterns are consistent with China's rapid industrialization and the challenge of decoupling growth from emissions. The model's high accuracy in this context suggests that traditional energy indicators remain key predictors and should remain focal points for emission reduction strategies in the near term.

In contrast, the United States features a distinct set of emission

drivers including oil consumption, electricity generation by fuel type, and transportation-related emissions. This reflects a more diversified energy portfolio and the significant impact of the mobility sector on national carbon output. The relatively lower Jaccard similarity with China underscores the need for context-specific emission modeling and policy tailoring.

The United Kingdom and France both emphasize electricity structure, such as the shares of coal, oil, and gas in generation, as well as indicators related to carbon emissions from flaring or industrial sectors. These reflect post-industrial transition patterns, with increasing reliance on renewable energy and emphasis on regulatory control. The shared core features and high intersection with China's feature set suggest that while economic structures diverge, monitorable indicators converge in global accounting systems, offering a foundation for international cooperation on climate reporting.

Importantly, the UpSet plot and Jaccard analysis confirm that while there is a universal core of 21 common features, a substantial portion of features remain region-specific. This validates the need for adaptive, data-driven modeling approaches that do not assume uniformity in emission drivers. By using a hybrid XGBoost model optimized via differential evolution, our approach not only improves accuracy but also enhances interpretability and flexibility, enabling it to be used as a policy-support tool.

From an environmental governance perspective, the findings suggest that machine learning models must be tightly coupled with domain knowledge to effectively guide mitigation strategies. This work contributes toward building regionally sensitive carbon forecasting tools that can support pathway design for carbon neutrality and inform nationally determined contributions.

5. Conclusion

The present paper proposes a novel hybrid deep learning approach for the prediction of carbon emissions. In this study, we evaluate the prediction performance of our method for seven regions: Africa, Asia and the Pacific, CIS, Europe, the Middle East, North America, South America and Central America. The main findings of this study, based on the experimental results, are as follows.

The prediction of carbon emissions is not solely influenced by energy-related factors. Other seemingly unrelated features also play a role. Feature engineering allows for the selection and grouping of features. This ensures the efficiency of feature utilization in prediction and improves prediction accuracy.

The integrated model proposed in this paper is adaptable, with its basis in the performance of different models. The model achieves optimal integration results and excels in multicriteria prediction.

However, a notable limitation lies in the manual selection of hyperparameters for base models, which may constrain performance and scalability. To address this, future work will explore automated hyperparameter optimization techniques, such as Bayesian optimization (Victoria and Maragatham, 2021), simulated annealing (Wang et al., 2001), differential evolution (Wu et al., 2023), and particle swarm optimization (Tang et al., 2016).

Among these, Bayesian optimization is particularly promising due to its ability to efficiently explore high-dimensional and non-convex parameter spaces using probabilistic surrogate models. By iteratively balancing exploration and exploitation, it can identify near-optimal hyperparameter configurations with fewer evaluations compared to grid or random search. The adoption of such techniques is expected to significantly enhance model generalization, improve prediction accuracy, and reduce computational costs. Moreover, automated tuning can increase model robustness to varying data distributions across regions and facilitate more consistent performance in real-world carbon emission forecasting scenarios.

CRediT authorship contribution statement

Lingfei Zhang: Writing – original draft, Methodology, Data curation. **Gang Lu:** Formal analysis, Data curation, Conceptualization. **Xiaoqing Yan:** Project administration, Investigation. **Peng Xia:** Validation, Supervision, Resources. **Zhong Chen:** Visualization, Supervision. **Di Wu:** Writing – review & editing, Supervision.

Software and data

Name of software: DEOHXGBOOST, Developers: Lingfei Zhang
Contact: brinca@email.swu.edu.cn.

Date first available: March 16, 2025.

Software required: Pycharm

Program language: Python.

Source code at: <https://github.com/lingfei0804/DEOHXGBOOST>.

Documentation: Detailed documentation with code and data can be found at <https://github.com/lingfei0804/DEOHXGBOOST/blob/master/README.md>.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Gang Lu reports a relationship with Science and Technology Foundation of State Grid Corporation of China that includes: funding grants. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported in part by the Science and Technology Foundation of State Grid Corporation of China under grant 1400-202357341A-1-1-ZN (Identification of Energy Security Risks and Strategic Path Optimization Technology Research under Global Coal-Oil-Gas-Electricity Coupling in China).

Data availability

Our datasets can be found at the following link: <https://github.com/lingfei0804/DEOHXGBOOST>

References

- Aboagye, P.D., Sharifi, A., 2023. Post-fifth assessment report urban climate planning: lessons from 278 urban climate action plans released from 2015 to 2022. *Urban Clim.* 49, 101550.
- Azam, M.H., Hasan, M.H., Hassan, S., Abdulkadir, S.J., 2020. Fuzzy type-1 triangular membership function approximation using fuzzy c-means. In: 2020 International Conference on Computational Intelligence (ICCI), pp. 115–120. IEEE.
- Chen, S., Li, Y., Lin, J.-H., Chang, C.-H., 2024. Energy transition: cap-and-trade and carbon capture and storage for achieving net-zero emissions with sustainable insurance. *J. Environ. Manag.* 366, 121902. <https://doi.org/10.1016/j.jenvman.2024.121902>. <https://www.sciencedirect.com/science/article/pii/S0301479724018887>.
- Cruz, Y.J., Rivas, M., Quiza, R., Villalonga, A., Haber, R.E., Beruvides, G., 2021. Ensemble of convolutional neural networks based on an evolutionary algorithm applied to an industrial welding process. *Comput. Ind.* 133, 103530. <https://doi.org/10.1016/j.compind.2021.103530>. <https://www.sciencedirect.com/science/article/pii/S0166361521001378>.
- De Myttenaere, A., Golden, B., Le Grand, B., Rossi, F., 2016. Mean absolute percentage error for regression models. *Neurocomputing* 192, 38–48.
- Ding, S., Xu, N., Ye, J., Zhou, W., Zhang, X., 2020. Estimating Chinese energy-related co₂ emissions by employing a novel discrete grey prediction model. *J. Clean. Prod.* 259, 120793.
- Dombi, J., 1990. Membership function as an evaluation. *Fuzzy Set Syst.* 35 (1), 1–21.
- Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q., 2020. A survey on ensemble learning. *Front. Comput. Sci.* 14, 241–258.
- Epskamp, S., Fried, E.I., 2018. A tutorial on regularized partial correlation networks. *Psychol. Methods* 23 (4), 617.

- Fisher, R., 1920. Accuracy of observation, a mathematical examination of the methods of determining, by the mean error and by the mean square error. *Mon. Not. Roy. Astron. Soc.* 80, 758–770, 758–770 80.
- Gan, N., Zhao, S., 2024. Global greenhouse gas reduction forecasting via machine learning model in the scenario of energy transition. *J. Environ. Manag.* 371, 123309. <https://doi.org/10.1016/j.jenvman.2024.123309>. <https://www.sciencedirect.com/science/article/pii/S030147972403295X>.
- Hameed, I.A., 2011. Using Gaussian membership functions for improving the reliability and robustness of students' evaluation systems. *Expert Syst. Appl.* 38 (6), 7135–7142.
- Hoffmann Souza, M.L., da Costa, C.A., de Oliveira Ramos, G., 2023. A machine-learning based data-oriented pipeline for prognosis and health management systems. *Comput. Ind.* 148, 103903. <https://doi.org/10.1016/j.compind.2023.103903>. <https://www.sciencedirect.com/science/article/pii/S0166361523000532>.
- Hosseini, H., 2021. A deep learning model for predicting river flood depth and extent. *Environ. Model. Software* 145, 105186.
- Hou, Y., Wang, Q., Tan, T., 2023. Regional suitability assessment for straw-based power generation: a machine learning approach. *Energy Strategy Rev.* 49, 101173.
- Koutroumanidis, T., Iliadis, L., Sylaios, G.K., 2006. Time-series modeling of fishery landings using arima models and fuzzy expected intervals software. *Environ. Model. Software* 21 (12), 1711–1721.
- Li, W.-k., Wen, H.-x., Nie, P.-y., 2023. Prediction of china's industrial carbon peak: based on gdim-mc model and lstm-nn model. *Energy Strategy Rev.* 50, 101240.
- Liu, B., Chen, S., Gao, L., 2025. Combining residual convolutional lstm with attention mechanisms for spatiotemporal forest cover prediction. *Environ. Model. Software* 183, 106260.
- Mallipeddi, R., Suganthan, P.N., Pan, Q.-K., Tasgetiren, M.F., 2011. Differential evolution algorithm with ensemble of parameters and mutation strategies. *Appl. Soft Comput.* 11 (2), 1679–1696.
- McLeod, A.I., 2005. Kendall rank correlation and mann-kendall trend test. *R package Kendall* 602, 1–10.
- Moe, S.J., Madsen, A.L., Connors, K.A., Rawlings, J.M., Belanger, S.E., Landis, W.G., Wolf, R., Lillicrap, A.D., 2020. Development of a hybrid Bayesian network model for predicting acute fish toxicity using multiple lines of evidence. *Environ. Model. Software* 126, 104655.
- Pathania, A., Gupta, V., 2025. Interpretable transformer model for national scale drought forecasting: Attention-driven insights across India. *Environ. Model. Software* 187, 106394.
- Rao, C., Zhang, Y., Wen, J., Xiao, X., Goh, M., 2023. Energy demand forecasting in China: a support vector regression-compositional data second exponential smoothing model. *Energy* 263, 125955.
- Ren, Z., Zheng, J., Jiao, L., He, M., Hou, X., Coffman, D., Wang, S., 2024. The synergic impacts of air pollution control policies on pollutants and carbon emissions. *J. Environ. Manag.* 370, 122730. <https://doi.org/10.1016/j.jenvman.2024.122730>. <https://www.sciencedirect.com/science/article/pii/S0301479724027166>.
- Sedgwick, P., 2012. Pearson's correlation coefficient. *Bmj* 345.
- Shao, Z., Gao, S., Zhou, K., Yang, S., 2024. A new multiregional carbon emissions forecasting model based on a multivariable information fusion mechanism and hybrid spatiotemporal graph convolution network. *J. Environ. Manag.* 352, 119976. <https://doi.org/10.1016/j.jenvman.2023.119976>. <https://www.sciencedirect.com/science/article/pii/S0301479723027640>.
- Shiwakoti, R.K., Charoenlarpnopparut, C., Chapagain, K., 2023. Time series analysis of electricity demand forecasting using seasonal arima and an exponential smoothing model. In: 2023 International Conference on Power and Renewable Energy Engineering (PREE). IEEE, pp. 131–137.
- Sime, D.M., Wang, G., Zeng, Z., Peng, B., 2023. Uncertainty-aware and dynamically-mixed pseudo-labels for semi-supervised defect segmentation. *Comput. Ind.* 152, 103995.
- Tang, D., Dai, M., Salido, M.A., Giret, A., 2016. Energy-efficient dynamic scheduling for a flexible flow shop using an improved particle swarm optimization. *Comput. Ind.* 81, 82–95. <https://doi.org/10.1016/j.compind.2015.10.001> emerging ICT concepts for smart, safe and sustainable industrial systems. <https://www.sciencedirect.com/science/article/pii/S0166361515300476>.
- Victoria, A.H., Maragathan, G., 2021. Automatic tuning of hyperparameters using Bayesian optimization. *Evolving Systems* 12 (1), 217–223.
- Wang, T., Wu, K., Liu, Y., 2001. A simulated annealing algorithm for facility layout problems under variable demand in cellular manufacturing systems. *Comput. Ind.* 46 (2), 181–188. [https://doi.org/10.1016/S0166-3615\(01\)00107-5](https://doi.org/10.1016/S0166-3615(01)00107-5). <https://www.sciencedirect.com/science/article/pii/S0166361501001075>.
- Woolson, R.F., 2007. Wilcoxon signed-rank Test. Wiley encyclopedia of clinical trials, pp. 1–3.
- Wu, D., Sun, B., Shang, M., 2023. Hyperparameter learning for deep learning-based recommender systems. *IEEE Transactions on Services Computing* 16 (4), 2699–2712.
- Wu, Q., Zhang, W., Xu, N., Chen, Y., Liu, J., Jiang, L., 2025. Empowering local governments: how environmental fiscal federalism affects greenhouse gas emissions in China? *J. Environ. Manag.* 376, 124449. <https://doi.org/10.1016/j.jenvman.2025.124449>. <https://www.sciencedirect.com/science/article/pii/S0301479725004256>.
- Xiao, C., Ye, J., Esteves, R.M., Rong, C., 2016. Using spearman's correlation coefficients for exploratory data analysis on big dataset. *Concurrency Comput. Pract. Ex.* 28 (14), 3866–3878.
- Xiao, M., Yang, B., Wang, S., Zhang, Z., Tang, X., Kang, L., 2022. A feature fusion enhanced multiscale cnn with attention mechanism for spot-welding surface appearance recognition. *Comput. Ind.* 135, 103583. <https://doi.org/10.1016/j.compind.2021.103583>. <https://www.sciencedirect.com/science/article/pii/S0166361521001901>.
- Zeng, A., Ho, H., Yu, Y., 2020. Prediction of building electricity usage using gaussian process regression. *J. Build. Eng.* 28, 101054.
- Zimmerman, D.W., Zumbo, B.D., 1993. Relative power of the wilcoxon test, the friedman test, and repeated-measures anova on ranks. *J. Exp. Educ.* 62 (1), 75–86.